DNA Diagnostics—Molecular Techniques and Automation

ULF LANDEGREN, ROBERT KAISER, C. THOMAS CASKEY, LEROY HOOD

Molecular biology has revolutionized the understanding of many aspects of human disease. Ongoing developments in DNA diagnostics—the analysis of disease at the nucleic acid level—will soon provide automated, rapid, and inexpensive analyses for DNA or RNA sequences associated with genetic, malignant, and infectious diseases. DNA diagnostics will also facilitate the identification of disease-associated genes at birth, thus creating new opportunities for preventive medicine.

ADVANCES IN MOLECULAR BIOLOGY AND BIOTECHNOLOGY are creating exciting possibilities for DNA diagnostics (*1*). These developments will profoundly alter many aspects of modern medicine, including the pre- or postnatal analysis of genetic diseases, the identification of individuals predisposed to conditions such as diabetes or coronary heart disease, the analysis of infectious disease ranging from common colds to AIDS, and the diagnosis and classification of malignant disease.

Fundamental to DNA diagnostics is (i) the definition of nucleic acid sequences that are informative in disease and (ii) the development of techniques that permit monitoring these sequences in individual patients. Both of these tasks are complicated by the vast dimensions of the genome (Fig. 1). The human genome contains about 10^5 genes, encoded by roughly 3 to 5% of the total 3×10^9 base pairs of DNA (*2*). This DNA sequence is distributed on 24 different chromosomes. Every individual inherits a complete set of 22 autosomes and 1 sex chromosome, X or Y, from each parent. Thus, each autosomal gene is present in two copies. Genes are divided into exons (coding regions) and introns (intervening DNA sequences) and individual genes may span up to 2 million base pairs.

The identification and analysis of nucleic acid sequences of diagnostic value in a wide variety of disease states proceeds at a rapid rate. In the future, the availability of the complete sequence of the human genome, analyzed with appropriate algorithms, should make it possible to identify virtually all human genes; the accompanying complete physical map of the human genome will serve as the ultimate source of DNA probes for any human gene (*3*). DNA sequences associated with disease may differ from corresponding normal sequences in single nucleotide positions as well as by grosser changes such as deletions, insertions, duplications, and translocations of DNA segments. In the following sections we will consider currently available DNA and RNA diagnostic techniques, some of the important applications for this technology, and the requirements for further technical advances and automation.

Drs. Landegren, Kaiser, and Hood are in the Division of Biology at the California Institute of Technology, Pasadena, CA 91125. Dr. Caskey is associated with the Howard Hughes Medical Institute for Molecular Genetics, Baylor College of Medicine, Houston, TX 77030.

Techniques for Analysis of Nucleic Acid Sequences

Nucleic acid sequence analysis is approached by a combination of (i) physicochemical techniques, based on the hybridization or denaturation of a probe strand plus its complementary target, and (ii) enzymatic reactions with nucleic acid–specific cleaving (endonucleases), joining (ligases), and synthesizing (polymerases) enzymes. Nucleic acids may be studied at the DNA or at the RNA level. The former analyzes the genetic potential of individual humans and the latter the expressed information of particular cells.

Detection of membrane-immobilized target sequences. In the versatile DNA (Southern) blot technique, DNA is digested with restriction enzymes that specifically cleave DNA molecules at their individual recognition sequences. The DNA fragments are separated according to size in an agarose gel and transferred (blotted) onto a nitrocellulose or nylon membrane (*4*). Conventional electrophoresis separates fragments ranging from 100 to 30,000 base pairs while pulsed field gel electrophoresis resolves fragments up to 20 million base pairs in length (*5, 6*) (Fig. 1). The location on the membrane of a restriction fragment containing a particular gene is determined by hybridization with a specific, labeled nucleic acid probe. A single nucleotide change can be identified if it disrupts or creates a restriction enzyme recognition sequence altering the size of a DNA fragment (Fig. 2).

Immobilized DNA sequences may be analyzed by probing with allele-specific oligonucleotide (ASO) probes (*7*), which are synthetic DNA oligomers of approximately 20 nucleotides. These probes are long enough to represent unique sequences in the genome, but sufficiently short to be destabilized by an internal mismatch in their hybridization to a target molecule. Thus, any sequences differing at single nucleotides may be distinguished by the different denaturation behaviors of hybrids between the ASO probe and normal or mutant targets under carefully controlled hybridization conditions (Fig. 2).

Detection of target sequences in solution. Several rapid techniques that do not require nucleic acid purification or immobilization have been developed. For example, probe/target hybrids may be selectively isolated on a solid matrix, such as hydroxylapatite, which preferentially binds double-stranded nucleic acids (*8*). Alternatively, probe nucleic acids may be immobilized on a solid support and used to capture target sequences from solution. Detection of the target sequences can be detected with the aid of a second, labeled probe that is either displaced from the support by the target sequence in a competition-type assay (*9*) or joined to the support via the bridging action of the target sequence in a sandwich-type format (*10*). In

general, single-nucleotide distinctions are not possible with these approaches, which have been most useful for the detection of infectious agents.

We have recently developed a strategy that shares with ASO hybridization the capability to distinguish DNA sequences differing by any single nucleotide substitution. In the oligonucleotide ligation assay (OLA), the enzyme DNA ligase is used to covalently join two synthetic oligonucleotide sequences selected so that they can base pair with a target sequence in exact head-to-tail juxtaposition (*11*). Ligation of the two oligomers is prevented by the presence of mismatched nucleotides at the junction region (Fig. 2). This procedure allows for the distinction between known sequence variants in samples of cells without the need for DNA purification. The joining of the two oligonucleotides may be monitored by immobilizing one of the two oligonucleotides and observing whether the second, labeled oligonucleotide is also captured. The technique should be suitable for automated genetic analysis.

Scanning techniques for detection of base substitutions. Three techniques permit the analysis of probe/target duplexes several hundred base pairs in length for unknown single-nucleotide substitutions or other sequence differences. In the ribonuclease (RNase) A technique, the enzyme cleaves a labeled RNA probe at positions where it is mismatched to a target RNA or DNA sequence (*12*). The fragments may be separated according to size and the approximate position of the mutation identified. In the denaturing gradient gel technique, a probe-target DNA duplex is analyzed by electrophoresis in a denaturing gradient of increasing strength. Denaturation is accompanied by a decrease in migration rate. A duplex with a mismatched base pair denatures more rapidly than a perfectly matched duplex (*13*). The RNase A and denaturing gel techniques each detect approximately 50% of all single base changes (Fig. 2). A third method relies on chemical cleavage of mismatched base pairs (*14*). A mismatch between T and C, G, or T, as well as mismatches between C and T, A, or C, can be detected in heteroduplexes. Reaction with osmium tetroxide (T and C mismatches) or hydroxylamine (C mismatches) followed by treatment with piperidine cleaves the probe at the appropriate mismatch.

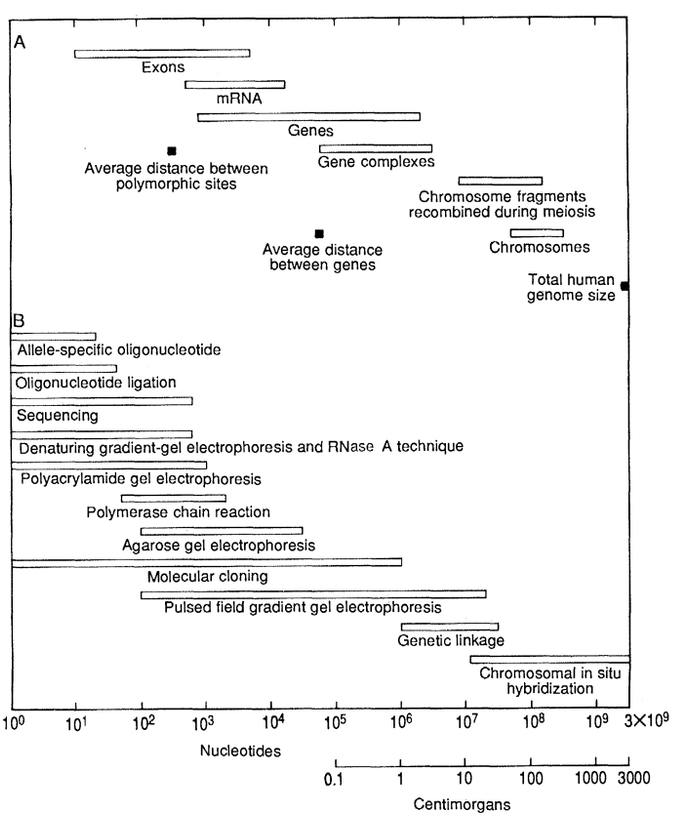

Fig. 1. Diagnostic techniques and the human genome. (**A**) The ranges of sizes of informational units of the human genome in nucleotide or centiMorgans. (**B**) The size ranges over which various diagnostic techniques are useful.

DNA sequence analysis. The ultimate tool for detecting DNA sequence variants is DNA sequence analysis. The sequence of the nucleotide residues along a single DNA strand may be determined over 300 to 600 nucleotides by the chemical degradation technique of Maxam and Gilbert (*15*) or the chain termination technique of Sanger *et al.* (*16*). In these techniques, four sets of DNA fragments terminating, respectively, at every G, A, T, or C across an unknown sequence are separated by electrophoresis in four lanes on a polyacrylamide gel. The gel has the capability to resolve DNA fragments differing in length by just a single nucleotide (Fig. 3). The size order of the rungs in the ladders of DNA fragments generated in these four reactions identifies the DNA sequence. Automated DNA sequencing (*17–19*) is likely to replace the scanning techniques described previously since it provides precise information about a mutation that can then be used in simpler tests to follow the inheritance of the mutation in an affected family.

Chromosomal in situ hybridization. The location of genes on chromosomes may be determined by microscopy after in situ hybridization of DNA probes to condensed metaphase chromosomes from dividing cells (*20*).

RNA detection. RNA molecules are analyzed either to assess the quantitative level of expression of particular genes or to identify structural mutations. Specific RNA molecules may be separated in gels according to size, transferred onto a membrane, and detected with a complementary probe (RNA or Northern blot) (*21*). In a more rapid approach, quantitative and structural information about individual RNA species may be acquired by solubilizing tissue in a denaturing solution that inactivates RNA-degrading enzymes but permits hybridization to a labeled probe in solution (*22, 23*). The protection of the probe by target molecules during a subsequent digestion with a single strand–specific enzyme is evaluated by gel electrophoresis (*12, 22, 24*). The identification of RNA molecules in a tissue section by means of in situ hybridization with a labeled probe (*25, 26*) has found important applications in molecular pathology. RNA molecules may also be converted into more stable DNA molecules by reverse transcription with an RNA-dependent DNA polymerase, and then studied by the techniques available for DNA analysis.

To use these analytical techniques, it is necessary to have a sensitive means of detecting the probe. A variety of reporter groups exist for the labeling of nucleic acid probes with detectable moieties.

Labeling of Nucleic Acid Probes

The most common nucleic acid labels are radioisotopes, primarily ^{32}P and ^{35}S, which are easily introduced by a variety of enzymatic techniques (*27*) (Table 1). They offer the possibility of incorporating many labels per probe, thus increasing the sensitivity of detection. Disadvantages of radioisotopes include the cumbersome nature of autoradiographic or scintillation-counting methods of detection, isotope instability, health hazards, and disposal problems.

Enzymes such as peroxidase and alkaline phosphatase may be chemically coupled to synthetic oligonucleotide probes (*28*), and offer the potential for high-sensitivity detection through the catalytic action of the enzyme on an appropriate substrate. The binding of enzyme to the probe may also be achieved by enzymatically or

chemically incorporating into the probe multiple biotin groups that may subsequently be bound, in a virtually irreversible affinity interaction, by conjugates of the proteins avidin or streptavidin with enzyme molecules. The enzyme action may be detected as chromophore deposition, or the release of colored, fluorescent, and luminescent species in solution. Detection sensitivities similar to those for radioisotopes may be achieved by complexing many enzyme molecules to each probe (*29*) after the hybridization.

Fluorophores represent an attractive class of reporter molecules since they are stable, directly detectable, and offer an opportunity for the simultaneous detection of multiple probes by means of different fluorophores with discriminable emission spectra. They may be introduced into probes chemically (*30*) or enzymatically (*19*). In general, the detection of simple, organic groups by fluorescence is not sufficiently sensitive for DNA diagnostic applications. In contrast, the fluorescence from chelates of rare earth metal ions (such as europium or terbium) may be detected at levels similar to those of radioisotopes (*31*). Due to the long fluorescence lifetime after excitation of the metal ions relative to most organic fluorescent species, detection may be delayed until much of the intrinsic background fluorescence has decayed to near zero without a significant reduction of the fluorescence from the rare-earth metals. Such time-resolved measurements greatly augment the signal over background. It also appears that, in contrast to conventional organic fluorophores, multiple rare earth metal chelates may be added to a probe with a proportional gain in fluorescence because of a lack of negative interference.

Typically 5 to 10 μg of genomic DNA (about 2×10^{-18} mol) are required for the radioisotopic detection of a sequence by techniques such as the DNA blot, ASO, and OLA. This amount of DNA is present in about 10^6 nucleated cells, which can be obtained from less than 1 ml of blood or, prenatally, from either a chorionic villus sampling at the 10th week of pregnancy or an amniocentesis performed in the 16th week. Vanishingly small amounts of DNA may be used, however, if the analytical reaction is preceded by target amplification with a recently developed technique described below.

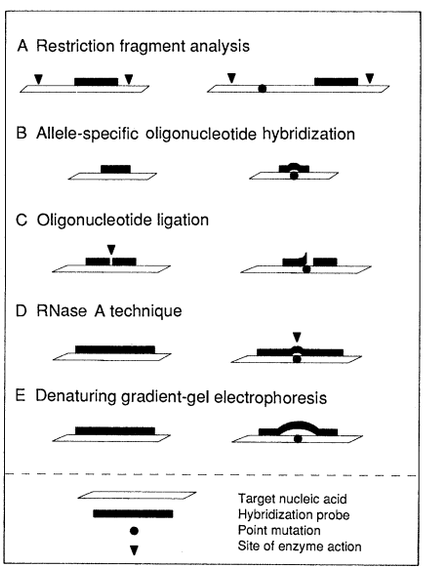

Fig. 2. Diagnostic techniques capable of distinguishing single-nucleotide substitutions. In each section, the diagram on the left represents the analysis of the normal target sequence and that on the right the analysis of a target sequence differing in one position.

Table 1. Reporter groups for labeling or nucleic acid probes.

Reporter*	Examples	Sensitivity of probe (10^{-18} mol)†	Method of detection
Radioisotope	^{32}P, ^{35}S, ^{125}I	0.05–2	Autoradiography, scintillation counting, gamma ray spectrometry
Enzyme	Peroxidase, alkaline phosphatase	0.2–1000	Colorimetry, fluorimetry, luminimetry
Fluorophore	Fluorescein, rhodamine	1–10,000	Fluorimetry
Chelated rare earth metal ion	Europium, terbium	1–100	Time-resolved fluorimetry

*Radioisotopes are generally incorporated into DNA probes by enzyme action (*27*), although chemical methods may also be used (*94*). Enzymes used as reporters can be coupled to nucleic acids either by direct chemical cross-linking to derivatized synthetic oligonucleotides (*28*) or by conjugation to avidin or streptavidin and affinity interaction of the conjugate with biotin-containing probe (*29*). Fluorescent organic molecules and metal ion chelates may be incorporated into DNA either by chemical synthesis (*30*) or by enzyme action on appropriately modified nucleoside triphosphates (*19*). †The observed sensitivity varies with the analytical procedure used in the assay, the instrumentation, and the number of reporter groups incorporated into each probe molecule.

DNA Amplification by the Polymerase Chain Reaction

A DNA segment of up to approximately 6000 base pairs in length may be amplified exponentially starting from as little as a single gene copy by means of the polymerase chain reaction (PCR) (*32, 33*). In this technique a denatured DNA sample is incubated with two oligonucleotide primers that direct the DNA polymerase–dependent synthesis of new complementary strands. Multiple cycles of synthesis each afford an approximate doubling of the amount of target sequence. Each cycle is controlled by simply varying the temperature to permit denaturation of the DNA strands, annealing the primers, and synthesizing new DNA strands. The use of a thermostable DNA polymerase obviates the necessity of adding new enzyme for each cycle, thus permitting fully automated DNA amplification. Twenty-five amplification cycles increase the amount of target sequence by approximately 10^6-fold. For the purpose of gene analysis the PCR technique offers the advantages of increased signal intensity in subsequent assays (*34*). The amplified DNA segment may be analyzed after immobilization for the presence of given sequence variants by means of ASO probes (*35*), in solution with the OLA technique (*11*), or with techniques that scan for base substitutions. Once sufficient quantities of the fragment are available, the nucleotide sequence can be determined. These techniques can be performed on samples of cells without prior DNA purification (*32*).

Automation of DNA Diagnostic Procedures

A number of instruments are available to facilitate routine DNA diagnostic procedures.

1) The DNA extractor prepares DNA suitable for DNA blots from eight tissue samples in 3 hours. This instrument purifies DNA by a cycle of proteolytic cell lysis, extraction of non-nucleic acid material with phenol and chloroform, ethanol precipitation, and filtration to collect the purified DNA. The primary advantages of the extractor are the ability to handle multiple samples simultaneously and the minimization of user manipulation. Manual methods of processing DNA frequently result in shearing and nuclease contamination. We expect the development of DNA extractors that are based on even simpler procedures and have the capacity to handle hundreds of samples daily.

2) The DNA synthesizer assembles oligonucleotides up to 200 nucleotides in length with a synthesis rate of 12 to 15 minutes per cycle (*36*). This machine utilizes a solid support–bound nucleoside

to initiate the sequential assembly of an oligomer in a 3′ to 5′ direction via activated nucleoside derivatives. Modified residues, including detectable and retrievable moieties, may be included in the synthesized strand. This instrument is critical for the synthesis of many DNA probes.

3) The DNA amplifier performs the PCR technique in 48 DNA samples simultaneously, amplifying individual DNA segments by 10^6-fold in 4 hours (*33*). It affords a simple and rapid means to greatly increase the signal-to-noise ratio in many DNA diagnostic assays.

4) Automation of a variety of routine manipulations including restriction enzyme analyses and DNA sequencing reactions is possible with laboratory robotic workstations. A robot arm, used in conjunction with a temperature-controlled, 96-well microtiter plate holder, completes 24 enzymatic DNA sequencing reactions in 45 minutes (*37*) (Fig. 4). It should be possible to combine the PCR technique with analysis by the OLA or by DNA sequencing in fully automated protocols.

5) We have developed an automated DNA sequencer (*17*) that uses the enzymatic sequencing technique (*16*) coupled with four-color fluorescence detection and computer-based analysis to decipher electrophoretic band patterns and determine DNA sequences (Fig. 3). A commercial version of this instrument can analyze 6000 nucleotides of sequence in 12 hours. Several other automated DNA sequencers have been described (*18, 19*). These instruments can also be adapted to the analysis and comparison of band patterns from restriction enzyme digests. We expect the sequencing capacity of these instruments to increase by at least tenfold in the next few years.

6) Several instruments exist that size-separate large DNA segments by subjecting the DNA molecules to electric fields of alternating orientation; this technique is known as pulsed field gradient gel electrophoresis (*5*). An improved version of these instruments has been developed that utilizes a hexagonal array of 24 computer-controlled electrodes to reproducibly generate defined, homogenous electric fields in agarose gels; it is known as the programmable autonomously-controlled electrodes (PACE) gel electrophoresis instrument (Fig. 5) (*6*). The orientation of multiple, distinct electric fields may be alternated during a single run in a preprogrammed fashion to allow the rapid, uniform separation of DNA fragments, thus optimizing resolution for sizes ranging from 500 base pairs to 10 million base pairs. Such an instrument, used in conjunction with fluorescence detection similar to that developed for the automated DNA sequencer, should be useful in identifying deletions, translocations, and amplifications of DNA sequences and in determining the distribution of restriction enzyme recognition sequences over large DNA segments.

7) Novel computing systems with specialized coprocessors, exemplified by the fast data-finder and the biological information signal processor, are being specially designed to handle the enormous computational tasks of searching for sequence similarity and sequence patterns in databases that will encompass billions of bytes of information (*38*) (Fig. 6). These processors will also be useful for combining data from instruments determining restriction maps and DNA sequences into a form suitable for direct entry into appropriate databases. New computing hardware and software will be an important component in the automation and large-scale analysis required for future DNA diagnostic instruments.

These techniques and instruments are rapidly being applied to human DNA diagnostics. We next discuss the technical challenges posed by the major applications for DNA diagnostics—genetic diseases, malignancy, infectious diseases, and forensic medicine.

Applications for DNA Diagnostic Techniques

Genetic disease. Single-gene disorders are grouped according to their mode of inheritance into (i) autosomal dominant diseases affecting individuals heterozygous for the mutant gene (one defective copy of the gene), (ii) autosomal recessive diseases affecting homozygous individuals (two defective gene copies), and (iii) X-linked recessive diseases predominantly affecting males with a mutation in the single (hemizygous) copy of an X-linked gene.

The defective gene product has been at least tentatively identified for 400 human monogenic disorders (*39*). The nature of the mutations involved has been determined in 45 of these disorders. Length mutations (insertions or deletions) were observed in 13 genetic diseases, single-nucleotide substitutions were observed in 23, and both types occurred in 9. The frequency and mode of inheritance of some common genetic disorders, based on two Canadian studies, are given in Table 2. The presence of a supernu-

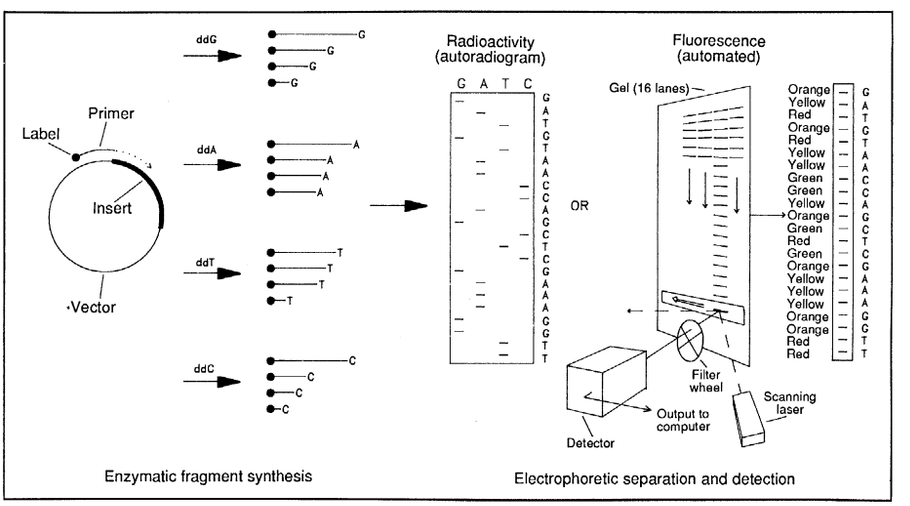

Fig. 3. DNA sequence analysis by chain termination. A synthetic oligonucleotide is used to prime the enzymatic synthesis of a complementary DNA strand for a single-stranded template (*16*). Four different strand extension reactions are performed in the presence of deoxynucleotides mixed with one of the chain-terminating nucleotide analogs (dideoxynucleotides ddG, ddA, ddT, or ddC). This produces four sets of nested reaction products, all beginning at the same point and ending at every G, A, T, and C residue, respectively, over a 600-nucleotide stretch. Radioactively labeled fragments are then separated according to size by electrophoresis in four adjacent lanes, one for each nucleotide, of a polyacrylamide gel. An image of the resolved fragments is obtained by autoradiography. The nucleotide sequence of the DNA segment is deduced from the size order of the resolved DNA fragments in the four different lanes. Alternatively, each of the four reactions may be color-coded by using sequencing primers labeled with a different fluorophore for each of the four sequencing reactions. The products of the reactions may then be combined and separated in a single lane of a polyacrylamide gel. The beam of an argon laser mechanically scans a horizontal segment at the bottom of the gel. As the fluorophore-labeled fragments undergoing electrophoresis pass through this region they are excited to fluorescence. The emitted light is collected and focused onto a detector after passing through four different filters on a rotating wheel. The identity of the terminating nucleotide analog of each passing fragment is identified because of the unique fluorophore associated with it. The temporal order of passing fluorescent molecules is automatically translated into linear DNA sequence and stored by the computer.

Table 2. Frequencies of common genetic disorders.

	Frequency* (per 10^6 births)	Inheritance	Chromosomal position identified†	Gene identified†	Mostly caused by new (n) or inherited (i) mutations‡
		Monogenic disorders			
Onset before age 25					
Cystic fibrosis	230	Recessive	Yes	No	i
Neurofibromatosis	85	Dominant	Yes	No	n and i
Muscular dystrophy	77	X-linked	Yes	Yes	n and i
Osteodystrophy	72	Dominant	Yes	Yes	n and i
Phenylketonuria	69	Recessive	Yes	Yes	i
Hemophilia A	56	X-linked	Yes	Yes	n and i
Tuberous sclerosis	50	Dominant	Yes	No	n and i
Onset after age 25‡					
Hypercholesterolemia	2000	Dominant	Yes	Yes	i
Otosclerosis	1000	Dominant	No	No	i
Polycystic kidney disease	800	Dominant	Yes	No	i
Huntington's disease	400	Dominant	Yes	No	i
Multiple polyposis coli	100	Dominant	Yes	No	I
		Chromosomal disorders			
Trisomy 21	1200				n
XXX	500				n
XXY	500				n
XYY	500				n
Trisomy 18	76				n

*Data have been described (*40, 48*). Frequencies of monogenic diseases vary widely between different populations. †The indicated disorders often include more than one clinical entity. Information given pertains to predominant forms (*95*). ‡These disorders mostly manifest after reproduction and are not strongly selected against.

merary (extra) chromosome represents a frequent genetic aberration (*40*), presently identified by microscopic evaluation of dividing cells. Thus Table 2 emphasizes the requirement for DNA detection techniques capable of distinguishing single nucleotide differences, as well as deleted or duplicated genetic material.

Many genetic diseases are caused by a limited number of mutant alleles or gene variants that are found at a high frequency in human populations. These kinds of divergent sequences are used to screen for carrier status and in fetal diagnosis. Examples include β thalassemia and sickle cell disease (*41*), α-l-antitrypsin deficiency disease (*42*), and phenylketonuria (*43*). Each of these diseases is caused predominantly by one or a few single-nucleotide differences. The high carrier frequency of some automosal recessive mutations in particular human populations can be explained by founder effects (prolific individuals pass the mutation to their progeny) or by selective advantages to heterozygous carriers that counterbalance the reduced fitness of individuals homozygous for the mutation (*44*).

Table 3. DNA diagnostic techniques used in conjunction with the PCR.

Application	Types of sequence detected	Diagnostic techniques	References
Lesch Nyhan disease	Single-nucleotide substitutions	PCR then DNA sequencing	(*55*)
Duchenne muscular dystrophy	Large deletions	PCR on multiple exons then ASO	(*54*)
Insulin-dependent diabetes	Multiple sequence variants	PCR then DNA sequencing or ASO	(*46*)
Carcinoma of the pancreas	Single nucleotide substitutions	PCR then ASO	(*73*)
Chronic myelogenous leukemia	Chromosomal translocation	PCR then DNA probe	(*83*)
AIDS	Presence of HIV genome	Reverse transcription then PCR then DNA probe	(*86*)
Forensics	Various polymorphisms	PCR from single hair then ASO	(*88*)

Combinations of genes encode complex aspects of human phenotype, such as the immune response and cholesterol metabolism. Defects in one or more of these genes can cause diseases that may be exacerbated by environmental factors. This group of polyfactorial diseases, exemplified by coronary artery disease (*45*), diabetes mellitus (*46*), and multiple sclerosis (*47*), is of great practical interest because it is far more common than the group of monogenic diseases (*48*). Furthermore, it may be possible for susceptible individuals to avoid known exogenous risk factors (such as diet or infectious agents) or watch for the development of symptoms that may then be treated at an early stage. For example, emphysema associated with α-l-antitrypsin deficiency is enhanced in frequency and severity by smoking, thus offering an obvious intervention possibility (*49*). Certain genes involved in cholesterol metabolism, such as variant forms of the low-density lipoprotein receptor (*45*) or the Apo B, C, and E genes (*50*), are predisposing factors in coronary artery disease. The highly polymorphic class II genes of the major histocompatibility complex are likely to have a role in several polyfactorial disorders (*46, 51*). The products of these genes participate in immune responses and are correlated with susceptibility to a large number of autoimmune diseases. Disease-associated class II alleles have traditionally been studied with antibodies to class II molecules. Recently, DNA sequence analysis of PCR-amplified class II gene segments has afforded increased precision in the definition of class II gene sequences associated with susceptibility to insulin-dependent diabetes mellitus and to the autoimmune skin disorder pemphigus vulgaris (*46, 52*).

X-linked recessive and autosomal dominant diseases that interfere with reproduction are maintained in the population largely because of new mutational events (Table 2). Since these cases occur in a sporadic fashion and the precise mutation cannot be predicted, genetic screening in all pregnancies is impractical. Genetic analysis in affected families, to identify the mutant gene in subsequent pregnancies, may be based either on the characterization of, and screening for, the mutant allele or on the study of linked genetic markers coinherited with the mutant gene. For some diseases caused by newly arising mutations, as exemplified by Duchenne muscular

dystrophy, the lesions are predominantly large deletions (*53*); in other cases, such as Lesch Nyhan disease, mostly single-nucleotide substitutions are seen (*23*). To detect such mutations, the techniques used must identify deletions or mutant nucleotide positions over extensive segments of DNA. Large deletions may be identified by pulsed field gel electrophoresis (*53*) or by amplifying and detecting individual exons (*54*). The mismatch scanning techniques may be used to identify single nucleotide substitutions (*12–14*). In an alternative approach, the hypoxanthine-guanine phosphoribosyltransferase gene (defective in Lesch Nyhan disease) was amplified from genomic DNA with PCR and the mutation was identified by DNA sequence analysis (*55*). This strategy permits the synthesis of specific DNA probes for study of individual families.

Whether or not the nature of a mutation or even the identity of the normal gene or protein is known, predictive testing in affected families may be carried out by following the inheritance of genetic markers linked to the mutant gene. This approach is presently being applied to diseases such as Duchenne muscular dystrophy, cystic fibrosis, adult polycystic kidney disease, and Huntington's disease (*56*).

The analysis of genetic linkage is based on the exchange of genetic information (recombination) between the pairs of maternally and paternally derived chromosome homologs during the meiotic cell divisions that give rise to the germ cells. The genetic distance between two markers is defined as 1 centiMorgan (cM) when the markers have a probability of 1 in 100 of recombining each generation. This corresponds to an average physical distance between the markers of 10^6 base pairs. The relation between the genetic and the physical maps is nonlinear due to the existence of preferential sites of recombination (*57*). The total genetic length of the human genome is estimated at 3300 cM, corresponding to an average of approximately three recombination sites per chromosome per generation (*58*).

Sequence polymorphisms, occurring once every few hundred nucleotides in the genome (*59*), are predominantly a consequence of single-nucleotide substitutions, sometimes changing the site of cleavage by restriction enzymes. Changes in the position of cleavage sites result in DNA fragments that exhibit size differences between individuals. These different forms are termed restriction fragment length polymorphisms (RFLPs) and represent valuable genetic linkage markers (*60*). One kind of RFLP is caused by the presence of variable numbers of tandemly repeated DNA sequences (VNTRs); VNTRs result in highly informative RFLPs present in multiple allelic forms (*61*).

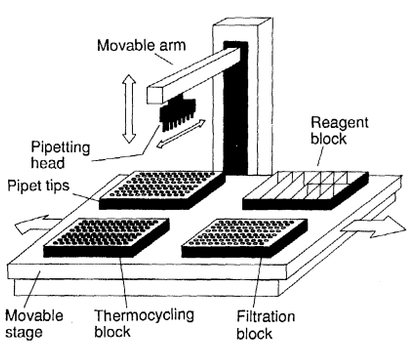

Fig. 4. Schematic diagram of an XYZ-type robotic workstation for the automation of DNA diagnostic techniques. This instrument has the capability to transfer reagents to and from 96-well microtiter plates. The stage holds pipette tips and reagent containers, as well as two blocks for the plates. The temperature-control block automatically adjusts and maintains the temperature of the wells in which reactions are performed, while the filtration block provides filtering capability in a 96-well format. The stage also houses a variety of pipetting heads (not shown) which may be automatically exchanged during operation. The stage allows movement in the X direction, while the arm and the pipetting head afford movement in the Z and Y directions, respectively.

Information about the genetic distance between two markers in linkage is obtained by analyzing the frequency of recombination between the markers in large families (*62*). A human linkage map with an average marker spacing of 10 cM has been established (*63*). This map offers a 95% chance of including any genetic locus. Improved computational strategies (*64*), along with denser linkage maps and alternative ways of analyzing genetic markers (*65*), will permit linkage analysis of an increasing number of human genetic diseases, including conditions with complex inheritance patterns.

The process of identifying a gene on the basis of its chromosomal location is known as reverse genetics (*66*). The successful use of genetic linkage techniques may localize the gene-containing region to within 1 to 10 cM. If the genome is estimated to include 10^5 genes, then on the average there would be 30 genes per centiMorgan. Identification of the disease gene depends on defining mutations within the candidate gene that correlate with the occurrence of the disease and ultimately demonstrating the direct link between the mutant gene and the impaired function. For Duchenne muscular dystrophy, chronic granulomatous disease, and retinoblastoma, reverse genetics has led to the successful identification of the relevant genes (*66, 67*). In each of these three conditions several different mutations, including large deletions, were available. These deletions simplified the isolation of the relevant gene segment. In contrast, the mutant genes causing Huntington's disease and cystic fibrosis, both likely to be caused by one or very few mutant alleles, have been mapped to particular chromosomal regions but have remained unidentified for a frustrating length of time (*68*). New techniques to identify mutant genes within large chromosomal regions are needed to accelerate reverse genetic approaches to disease gene identification.

Malignant disease. Many forms of human malignancy occur in familial as well as sporadic forms (*69*). Discrete genetic changes, characteristic of different types of neoplasm, have been observed frequently, and are thought to initiate or cause progression of the malignant phenotype. The identification of such changes represent the major diagnostic challenge in malignancy. Both dominant and recessive modes of neoplastic transformation have been observed (*70*). Dominant-acting oncogenes constitute a group of genes, totaling 40 at the present time (*70*), which can induce transformation when introduced into a cell or an animal in a structurally altered form or when improperly expressed. In recessive oncogenesis, tumor cells have lost both copies of an "anti-oncogene" (*71*) and, accordingly, the introduction of one normal chromosome would be expected to correct the phenotype.

Dominant-acting oncogenes may be activated by a variety of mutational mechanisms, exemplified below.

1) Single-nucleotide substitutions, modifying codon 12 of the c–K-*ras* gene are observed in a majority of pancreatic carcinomas and in approximately one third of all colorectal carcinomas (*72, 73*). The mutations alter the structure of the c–K-*ras* protein, which is involved in signal transmission at the cell membrane (*74*).

2) The Philadelphia chromosome, characteristic of chronic myelogenous leukemia, is the result of a chromosomal translocation. This translocation activates an oncogene, c-*abl*, through the substitution of its 5′ end by that of another gene, *bcr* (*75*). The breakpoints of the reciprocal translocation are spread over a 100,000–base pair region in *bcr* and a 6000–base pair region in c-*abl* in chronic myelogenous leukemia cells from different individuals. Still, the fusion proteins are identical except for the presence or absence of 24 amino acids encoded by one of the *bcr* exons. The *bcr/abl* fusion protein has a substantially increased tyrosine kinase activity, as compared to normal c-*abl*.

3) N-*myc* is amplified to 300 copies in some neuroblastomas, with concomitant increase of mRNA levels. This amplification is

strongly correlated with a poor prognosis (*76*). The amplification is frequently observed by molecular genetic techniques in the absence of such microscopic evidence for gene amplification as expanded chromosomal regions on specific chromosomes or extrachromosomal genetic elements (*77*).

Retinoblastoma, arising in the retinas of young children, serves as the paradigm for a tumor caused by a recessive-acting oncogene. Hereditary and spontaneous cases can be explained genetically by the presence of a mutation disrupting each of the copies of a gene at a locus Rb (*78*). In the familial form, one of the mutant gene copies is inherited. Loss of the second copy is often the consequence of incorrect segregation of chromosomes at mitosis or results from a mitotic recombination event (*79*). Such large-scale deletions can be detected as a loss of heterozygosity for linked genetic markers in tumor tissue.

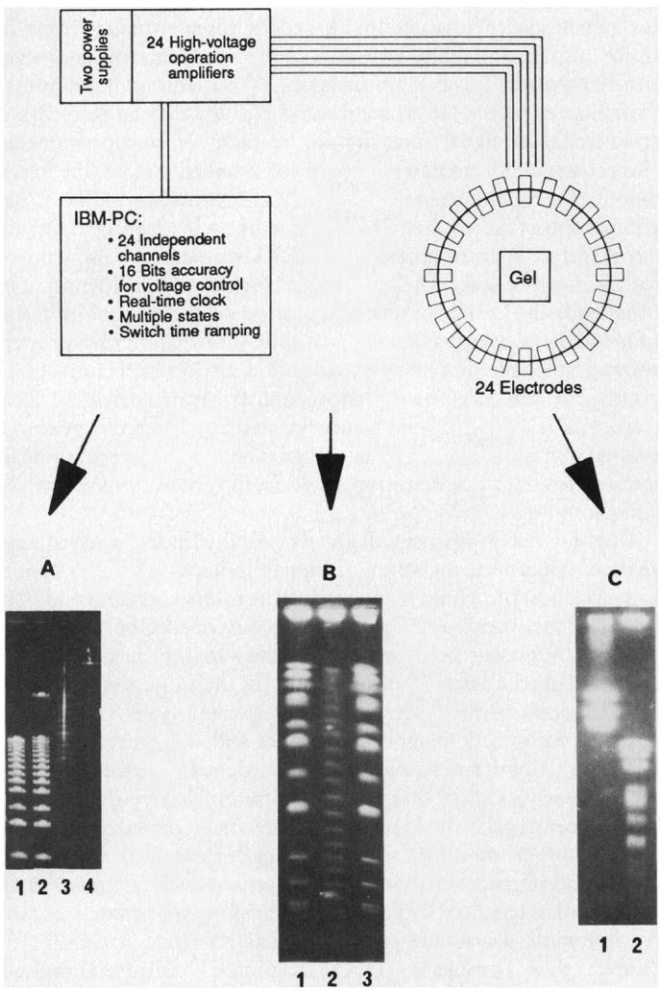

Fig. 5. Schematic diagram of the programmable autonomously controlled electrodes (PACE) gel electrophoresis apparatus (*6*). A set of 24 electrodes is arranged around an agarose gel in a closed contour. Each electrode is individually controlled by a high-voltage operation amplifier connected to a power supply. The 24 operation amplifiers are individually controlled by a computer, so that the voltage at each electrode may be accurately and reproducibly defined. Computer control allows for pre-programming of multiple voltage states and switch-time ramping during electrophoresis. By adjusting these two parameters, a nearly linear separation of DNA fragments can be obtained for any desired size range from 500 base pairs to 10 million base pairs. (**A**) Size range 1 to 50 kilobase pairs (kb): (lane 1) 1-kb ladder; (lane 2) 1-kb ladder plus 33.5-kb fragment; (lane 3) high molecular weight markers (BRL); (lane 4) λ DNA. (**B**) Size range 50 to 1000 kb: (lanes 1 and 3) chromosomes from the yeast *Saccharomyces cerevisae*; (lane 2) λ DNA concatamers. (**C**) Size range 200 to 7000 kb: (lane 1) chromosomes from the yeast *S. pombe*; (lane 2) *S. cerevisae* chromosomes.

Loss of heterozygosity for given chromosomal segments in similar tumors from different individuals may identify the chromosomal map position of suspected recessive oncogenes (*79, 80*). Such indications of a recessive etiology are observed for common solid tissue tumors such as carcinomas of the bladder, colon, breast, and lung (*81*).

Efficient assay techniques are needed to analyze tumors for RNAs of diagnostic value in malignancy, using batteries of probes for dominant- and recessive-acting oncogenes. Other gene probes are also of interest, as exemplified by the P-glycoprotein gene, which is specifically amplified in tumors and tumor cell lines resistant to multiple chemotherapeutic drugs (*82*). The PCR technique can be used to identify genotypic alterations from fine-needle biopsies, paraffinized histological sections (*73*), and sloughed cells (such as those from the urinary bladder). It also permits the detection of a low frequency of malignant cells in the bloodstream or the bone marrow in diseases such as chronic myelogenous leukemia when unique DNA or RNA segments are generated by the rearrangement process (*83*). Due to the ease with which DNA probes may be synthesized, it is possible to custom-design reagents, based on DNA sequence analysis, to detect recurrence or dissemination of malignant cells with characteristic genetic alterations.

Infectious diseases. Globally, infectious diseases represent a major threat to human health and nucleic acid analytical techniques offer improved diagnostic possibilities. A number of kits for detection of microbial nucleic acids in patient samples or in contaminated food have been approved by the U.S. Food and Drug Administration (*8, 84*). Detection of infectious agents at the level of DNA or RNA sequences has advantages over immunological detection techniques. For example, reagents may be designed to be specific for an individual organism or for a wide range of related organisms. Such reagents may be used to directly assay for microbial virulence factors and genes conferring resistance to antibiotics, which will aid in the rapid selection of the appropriate treatment. It will be necessary to develop even simpler, rapid assay formats that may be scored visually for bedside diagnosis.

While, for some purposes, techniques of moderate sensitivity will suffice, the detection of, for example, the HIV genome in infected

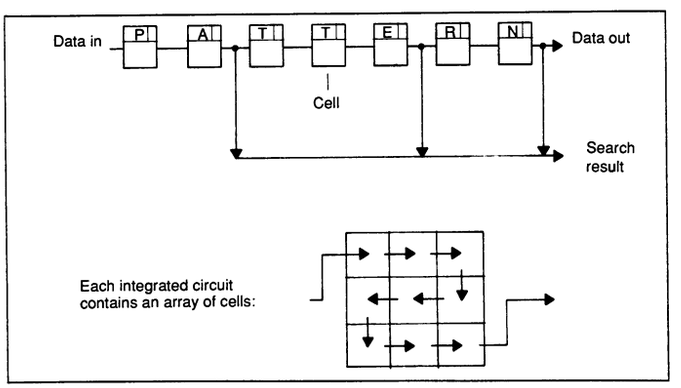

Fig. 6. Illustration of the fast data finder coprocessor for pattern matching. As depicted in the upper diagram, it is composed of a linear array of identical, specialized processing cells. Each cell is assigned a discrete operation to be performed on the pattern (or sequence) of interest. Data to be searched are then streamed through the array of cells much the same as water is through a pipe. Each cell operates on the data independently, communicating the results of its operation only to its immediately downstream neighbor and receiving additional information only from its immediately upstream neighbor. In this way, the accumulated degree of pattern matching follows the data through the pipeline. If the degree of comparison exceeds a previously defined threshold, this information is passed out of the fast data finder array to the host computer for further processing. Such arrays allow for the rapid execution of large numbers of concurrent operations.

blood may require finding sequences present as 1 part per 10 to 100,000 human genome equivalents (*26*). Contamination of blood banks with HIV, HTLV-I, hepatitis B, and hepatitis non-A non-B represents a serious threat to public health. With the current screening of blood for an antibody response to HIV, an estimated 460 recipients of transfused blood may become accidentally infected each year, mostly because the diagnostic antibody response may take 8 weeks to develop in an infected blood donor (*85*). The HIV genome can be detected in a highly sensitive manner by the PCR either as an RNA molecule in unintegrated virus or as a DNA molecule representing the integrated provirus (*86*).

Forensics. The identification of individuals at the level of DNA sequence variation offers a number of practical advantages over such conventional criteria as fingerprints, blood type, or physical characteristics. In contrast to most phenotypic markers, DNA analysis readily permits the deduction of relatedness between individuals such as is required in paternity testing. Genetic analysis has proven highly useful in bone marrow transplantation, where it is necessary to distinguish between closely related donor and recipient cells. Two types of probes are now in use for DNA fingerprinting by DNA blots. Polymorphic minisatellite DNA probes identify multiple DNA sequences, each present in variable forms in different individuals, thus generating patterns that are complex and highly variable between individuals (*87*). VNTR probes identify single sequences in the genome, but these sequences may be present in up to 30 different forms in the human population as distinguished by the size of the identified fragments (*61*). The probability that unrelated individuals will have identical hybridization patterns for multiple VNTR or minisatellite probes is very low. Much less tissue than that required for DNA blots, even single hairs, provides sufficient DNA for a PCR-based analysis of genetic markers (*88*). Also, partially degraded tissue may be used for analysis since only small DNA fragments are needed. We believe forensic DNA analyses will eventually be carried out with polymorphic DNA sequences that can be studied by simple automatable assays such as OLA. For example, the analysis of 22 separate gene sequences, each one present in two different forms in the population, could generate 10^{10} different outcomes, permitting the unique identification of human individuals. Such markers offer the convenience of a bar code for the analysis, storage, and comparison of test results obtained from large human populations.

Future Directions

The wide-scale application of DNA diagnostics will depend on the development of routine nucleic acid sequence detection procedures that are accurate, sensitive, rapid, and economical. Amplification of DNA by the PCR method in conjunction with an appropriate analytic technique provides a sensitive assay with enhanced accuracy that may be fully automated in the future. A number of different analyses have already been conveniently combined with amplification by the PCR (Table 3).

With the development of improved reporter groups (*31*) amplification will not be required for some analyses. This would be an advantage in situations where the PCR technique may introduce a source of error, such as when quantitation is critical or when numerous genetic loci are studied in a single sample. The possibility of performing multiple simultaneous analyses in a single nucleic acid sample will be valuable in genetic screening, in the investigation of genetic markers in linkage analysis, for the detection of infectious agents, and also in surveying patterns of RNA expression in tumor tissue. Potentially, fetal trophoblast cells could be isolated from maternal peripheral blood by the fluorescence-activated cell sorter to provide easy access to samples of the fetal genome for genetic screening (*89*).

For signal detection, radioisotopes will have to give way to stable, harmless reporter moieties that are immediately detectable. By suitably combining the described analytical techniques it will be possible to develop detection procedures that will be automated from the introduction of tissue samples to the recording and interpretation of the result.

The vast majority of routine DNA diagnostic problems may be addressed by just a few standard techniques. It is probable that instruments that detect and quantitate defined DNA and RNA sequences with large sets of ready-to-use probes will become standard equipment in clinical settings. Many other nucleic acid–analytical techniques will have predominantly research applications and will be useful for the identification of new probe sequences. Examples of possible clinical uses for DNA sequence detection techniques, such as PCR combined with OLA, include the screening for variant alleles associated with genetic disease, tissue-typing for major histocompatibility genes in transplantation, and analysis of these genes and T-cell receptor genes (*90*) in autoimmune disease. Complete or partial (*40*) chromosomal trisomy may be detectable as an increase in signal intensity for markers on a supernumerary chromosome. For prenatal diagnosis of X-linked disease, the inheritance of the mutant gene from a female carrier may be indicated by a comparison of linked X-chromosome markers in an affected son and a male fetus. Detection of defined DNA sequences is also valuable for diagnosis of infectious disease and for the unique identification of individuals. In the context of forensics, genetic analysis of a spot of blood could eventually provide helpful information regarding the sex and other physical characteristics of a suspect or victim. Analogously, the detection, quantitation, and structural analysis of RNA molecules is of great importance in characterizing the genes expressed in a tissue, primarily in malignancy. Similar applications for nucleic acid sequence detection also exist in veterinary medicine and agriculture.

Ongoing developments will greatly simplify nucleic acid sequence analysis. Moreover, the study of human molecular genetics continually extends the clinical applications for DNA diagnostics. The power of these advances is primarily a consequence of two circumstances: (i) nucleic acids lend themselves to the development of rapid and precise assay procedures, and (ii) the sequences monitored in the various diseases often have a direct causal role. The growing potential for genetic diagnosis contrasts with the currently limited therapeutic options in many genetic disorders. In the future we may expect to see a variety of approaches, exemplified by the introduction of normal genes into somatic cells to correct genetic defects (*91*) and the use of modified synthetic oligonucleotides for sequence-specific interference with viral gene function (*92*). The increased use of molecular genetic techniques in medicine will be associated with psychological, social, and ethical issues that require attention (*93*). Despite these problems, DNA diagnostics affords a radically changed perspective on human disease that will have consequences not only for diagnostic speed and accuracy but also for the development of novel therapeutic strategies.

REFERENCES AND NOTES

1. C. T. Caskey, *Science* **236**, 1223 (1987); B. Childs *et al.*, Eds., *Molecular Genetics in Medicine* (Elsevier, New York, 1988).
2. S. J. O'Brien, *Nature New Biol.* **242**, 52 (1973); J. O. Bishop, *Cell* **2**, 81 (1974).
3. L. Smith and L. Hood, *Bio/Technology* **5**, 933 (1987).
4. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975); J. Meinkoth and G. Wahl, *Anal. Biochem.* **138**, 267 (1984).
5. D. C. Schwartz and C. R. Cantor, *Cell* **37**, 67 (1984); G. F. Carle and M. V. Olsen, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3756 (1985); G. Chu, D. Vollrath, R. Davis, *Science* **234**, 1582 (1986).
6. S. M. Clark, E. Lai, B. W. Birren, L. Hood, *Science* **241**, 1203 (1988).
7. B. J. Conner *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 278 (1983).

8. D. E. Kohne, *Am. Clin. Prod. Rev.* **5**, 20 (1986).
9. C. P. H. Vary *et al.*, *Clin. Chem.* **32**, 1696 (1986).
10. M. Ranki *et al.*, *Gene* **21**, 77 (1983).
11. U. Landegren, R. Kaiser, J. Sanders, L. Hood, *Science* **241**, 1077 (1988).
12. R. M. Meyers *et al.*, *ibid.* **230**, 1242 (1985); E. Winter, F. Yamamoto, C. Almoguera, M. Perucho, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7575 (1985).
13. R. M. Meyers, N. Lumelsky, L. S. Lerman, T. Maniatis, *Nature* **313**, 495 (1985); R. M. Meyers, T. Maniatis, L. S. Lerman, *Methods Enzymol.* **155**, 501 (1987).
14. R. G. H. Cotton *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4397 (1988).
15. A. M. Maxam and W. Gilbert, *ibid.* **74**, 560 (1977); G. M. Church and S. Kieffer-Higgins, *Science* **240**, 185 (1988).
16. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
17. L. M. Smith *et al.*, *Nature* **321**, 674 (1986).
18. W. Ansorge, B. Sproat, J. Stegemann, C. Schwager, M. Zenke, *Nucleic Acids Res.* **15**, 4593 (1987); C. Connell *et al.*, *BioTechniques* **5**, 342 (1987); H. Kambara *et al.*, *Bio/Technology* 6, 816 (1988); J. A. Brumbaugh *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5610 (1988); J. K. Elder, D. K. Green, E. M. Southern, *Nucleic Acids Res.* **14**, 417 (1986); W. J. Martin *et al.*, *Bio/Technology* **3**, 911 (1985); A. Wada, M. Yamamoto, E. Soeda, *Rev. Sci. Instrum.* **54**, 1569 (1983); R. Staden, *Nucleic Acids Res.* **10**, 4731 (1982).
19. J. Prober *et al.*, *Science* **238**, 336 (1987).
20. B. Bhatt, J. Burns, D. Flannery, J. O'D. McGee, *Nucleic Acids Res.* **16**, 3951 (1988); H. C. MacGregor and S. Mizuno, *Chromosoma (Berl.)* **54**, 15 (1976).
21. P. S. Thomas, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5201 (1980).
22. J. Thompson and D. Gillespie, *Anal. Biochem.* **163**, 281 (1987); M. G. Pellegrino *et al.*, *BioTechniques* **5**, 452 (1987).
23. R. A. Gibbs and C. T. Caskey, *Science* **236**, 303 (1987).
24. A. J. Berk and P. A. Sharp, *Cell* **12**, 721 (1977).
25. A. Haase, M. Brahic, L. Stowring, H. Blum, *Methods Virol.* 7, 189 (1984); J. Bresser and M. J. Evinger-Hodges, *Gene Anal. Technol.* **4**, 89 (1987).
26. M. E. Harper *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 772 (1986).
27. B. Perbal, *A Practical Guide to Molecular Cloning* (Wiley, New York, 1988).
28. E. Jablonski, E. W. Moomaw, R. H. Tullis, J. L. Ruth, *Nucleic Acids Res.* **14**, 6115 (1986); B. C. F. Chu and L. E. Orgel, *ibid.* **16**, 3671 (1988); M. S. Urdea *et al.*, *ibid.*, p. 4937.
29. J. J. Leary, D. J. Brigati, D. C. Ward, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4045 (1983); E. L. Sheldon *et al.*, *ibid.* **83**, 9085 (1986).
30. J. Ruth, *DNA* **3**, 122 (1984); L. M. Smith, S. Fung, M. W. Hunkapiller, T. Hunkapiller, L. E. Hood, *Nucleic Acids Res.* **13**, 2399 (1985); B. A. Connolly and P. Rider, *ibid.*, p. 4485; B. A. Connolly, *ibid.* **15**, 3131 (1987); T. Kristensen, H. Voss, W. Ansorge, *ibid.*, p. 5507; J. Haralambidis *et al.*, *ibid.*, p. 4857.
31. E. Soini and H. Kojola, *Clin. Chem.* **29**, 65 (1983); A. Oser, W. K. Roth, G. Valet, *Nucleic Acids Res.* **16**, 1181 (1988); E. P. Diamandis, *Clin. Biochem.* **21**, 139 (1988).
32. R. K. Saiki *et al.*, *Science* **230**, 1350 (1985); R. K. Saiki, T. L. Bugawan, G. T. Horn, K. B. Mullis, H. A. Erlich, *Nature* **324**, 163 (1986).
33. R. K. Saiki *et al.*, *Science* **239**, 487 (1988).
34. The polymerase used for the PCR technique, Taq, has an accumulated error rate of one nucleotide incorrectly incorporated in 400 after 30 amplification cycles (*33*). This, however, generally does not affect analyses done on amplified molecules initially present in many copies, since the errors occur in a random fashion and each sequence variant will be present at a low frequency at any stage of the amplification process.
35. D. R. Engelke *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 544 (1988); G. McMahon, E. Davis, G. N. Wogan, *ibid.* **84**, 4974 (1987); C. Wong *et al.*, *Nature* **330**, 384 (1987); L. A. Wrischnik *et al.*, *Nucleic Acids Res.* **15**, 529 (1987).
36. M. H. Caruthers, *Science* **230**, 281 (1985); S. J. Horvath, J. R. Firca, T. Hunkapiller, M. W. Hunkapiller, L. Hood, *Methods Enzymol.* **154**, 314 (1987).
37. R. Wilson *et al.*, *BioTechniques* 6, 776 (1988).
38. D. P. Lopresti, *Computer* **20**, 98 (1987); Y. Kwang-I, S.-P. Hsu, R. E. Heiss, Jr., L. E. Hasiuk, *Pipelined for Speed: The Fast Data Finder System* (TRW Electronics and Defense Sector QUEST, Winter 1986/1987) pp. 5–17; S. Craig and K. Brewster, *Communications of ACM* **29**, 1229 (1986); M. Waterman, J. Peterson, T. Hunkapiller, personal communication.
39. V. A. McKusick, *Molecular Defects in Mendelian Disorders*, newsletter obtainable from Dr. McKusick, 25 March 1988.
40. T. J. Hassold and P. A. Jacobs, *Annu. Rev. Genet.* **18**, 69 (1984); L. Crippa, J. Ballaman, E. Engel, *Ann. Génét.* **27**, 190 (1984).
41. S. H. Orkin and H. H. Kazazian, Jr., *Annu. Rev. Genet.* **18**, 131 (1984).
42. V. J. Kidd, R. B. Wallace, K. Itakura, S. L. C. Woo, *Nature* **304**, 230 (1983).
43. A. G. DiLella and S. L. C. Woo, *Mol. Biol. Med.* **4**, 183 (1987).
44. J. I. Rotter and J. M. Diamond, *Nature* **329**, 289 (1987).
45. M. S. Brown and J. L. Goldstein, *Science* **232**, 34 (1986).
46. J. A. Todd, J. I. Bell, H. O. McDevitt, *Nature* **329**, 599 (1987).
47. A. D. Sadovnick and P. A. Baird, *Neurology* **38**, 990 (1988).
48. P. A. Baird *et al.*, *Am. J. Hum. Genet.* **42**, 677 (1988); B. K. Trimble and M. E. Smith, *Can. J. Genet. Cytol.* **19**, 375 (1977).
49. A. Janoff, *Am. Rev. Resp. Dis.* **132**, 417 (1985).
50. C. C. Shoulders and F. E. Baralle, *Methods Enzymol.* **128**, 727 (1986); J. Davignon, R. E. Gregg, C. F. Sing, *Arteriosclerosis* **8**, 1 (1988); L. Chan *et al.*, *Biochem. Biophys. Res. Commun.* **133**, 248 (1985); A. Law *et al.*, *Lancet* **i**, 1301 (1986).
51. Svejgaard, P. Platz, L. P. Ryder, *Immunol. Rev.* **70**, 193 (1983); J. A. Todd *et al.*, *Science* **240**, 1003 (1988).
52. A. A. Sinha *et al.*, *Science* **239**, 1026 (1988).
53. J. T. den Dunnen *et al.*, *Nature* **329**, 640 (1987).
54. J. S. Chamberlain *et al.*, in preparation.
55. R. A. Gibbs, P. N. Nguyen, C. T. Caskey, in preparation.
56. J. E. Spence *et al.*, *Hum. Genet.* 76, 5 (1987); G. J. Meissen *et al.*, *N. Engl. J. Med.* **318**, 535 (1988); S. T. Reeders *et al.*, *Nature* **317**, 542 (1985).
57. J. A. Kobori, E. Strauss, K. Minard, L. Hood, *Science* **234**, 173 (1986); M. Steinmetz, D. Stephan, K. F. Lindahl, *Cell* **44**, 895 (1986).
58. J. H. Renwick, *Br. Med. Bull.* **25**, 65 (1969); F. Vogel and A. G. Motulsky, *Human Genetics: Problems and Approaches* (Springer, Berlin, ed. 2, 1986).
59. D. N. Cooper, B. A. Smith, H. J. Cooke, S. Niemann, J. Schmidtke, *Hum. Genet.* **69**, 201 (1985).
60. Y. W. Kan and A. M. Dozy, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5631 (1978); D. Botstein, R. L. White, M. Skolnick, R. W. Davis, *Am. J. Hum. Genet.* **32**, 314 (1980).
61. Y. Nakamura *et al.*, *Science* **235**, 1616 (1987).
62. R. White and J.-M. Lalouel, in *Adv. Hum. Genet.* **16**, 121 (1987).
63. H. Donis-Keller *et al.*, *Cell* **51**, 319 (1987).
64. G. M. Lathrop, J. M. Lalouel, C. Julier, J. Ott, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3443 (1984); E. S. Lander and D. Botstein, *ibid.* **83**, 7353 (1987); E. S. Lander and D. Botstein, *Science* **236**, 1567 (1987); E. S. Lander and P. Green, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2363 (1987).
65. RFLP-based linkage analysis requires substantial amounts of DNA, often necessitating the establishment of permanent cell lines. Such analysis is also technically complex and not suitable for automation. However, genetic linkage markers (excluding VNTRs) may also be identified on PCR-amplified DNA fragments by, for example, the OLA, which requires little DNA and simple manipulations. Suitable sequence-scanning techniques such as the denaturing gradient technique may be applied to pooled human DNA to identify polymorphisms as well as their approximate allelic frequencies. Synthesis of amplification and detection reagents to monitor these markers can be done after the sequences of these allelic forms have been analyzed. The genetic map order of such markers could be determined by a novel strategy (Li *et al.*, *Nature*, in press). Markers from two or more genetic loci are amplified by the PCR starting from individual sperm cells that were derived from a donor heterozygous for both markers. Association in individual sperm of allelic forms of the genetic markers indicate genetic linkage. This method gives access to a virtually unlimited number of recombined chromosomes in sperm cells, hemizygous for all chromosomes, and could permit the development of a linkage map with a marker spacing of 1 cM.
66. S. H. Orkin, *Cell* **47**, 845 (1986).
67. A. Monaco *et al.*, *Nature* **323**, 646 (1986); B. Royer-Pokora *et al.*, *ibid.* **322**, 32 (1986); S. H. Friend *et al.*, *ibid.* **323**, 643 (1986); Y.-K. T. Fung *et al.*, *Science* **236**, 1657 (1987).
68. J. F. Gusella *et al.*, *Nature* **306**, 234 (1983); R. G. Knowlton *et al.*, *ibid.* **318**, 380 (1985); R. White *et al.*, *ibid.*, p. 382; B. J. Wainwright *et al.*, *ibid.*, p. 384.
69. A. G. Knudson, Jr., *Annu. Rev. Genet.* **20**, 231 (1986).
70. J. M. Bishop, *Science* **235**, 305 (1987); G. Klein, *ibid.* **238**, 1539 (1987).
71. E. J. Stanbridge, *Nature* **260**, 17 (1976); B. E. Weissman *et al.*, *Science* **236**, 175 (1987).
72. J. L. Bos *et al.*, *Nature* **327**, 293 (1987); K. Forrester, C. Almoguera, K. Han, W. E. Grizzle, M. Perucho, *ibid.*, p. 298.
73. C. Almoguera *et al.*, *Cell* **53**, 549 (1988).
74. M. C. Willingham, I. Pastan, T. Y. Shih, E. M. Scolnick, *ibid.* **19**, 1005 (1980); A. G. Gilman, *ibid.* **36**, 577 (1984).
75. S. Cory, in *Advances in Cancer Research*, G. Klein and S. Weinhous, Eds. (Academic Press, New York, 1986), p. 189.
76. R. C. Seeger, *N. Engl. J. Med.* **313**, 111 (1985).
77. G. M. Brodeur, R. C. Seeger, M. Schwab, H. E. Varmus, J. M. Bishop, *Science* **224**, 1121 (1984).
78. A. G. Knudson, Jr., *Proc. Natl. Acad. Sci. U.S.A.* **68**, 820 (1971).
79. W. K. Cavenee *et al.*, *Nature* **305**, 779 (1983).
80. H. J. Scrable, D. P. Witte, B. C. Lampkin, W. K. Cavenee, *ibid.* **329**, 645 (1987).
81. E. R. Fearon *et al.*, *ibid.* **318**, 377 (1985); E. Solomon *et al.*, *ibid.* **328**, 616 (1987); I. U. Ali, R. Lidereau, C. Theillet, R. Callahan, *Science* **238**, 185 (1987); K. Kok *et al.*, *Nature* **330**, 578 (1987).
82. J. R. Riordan *et al.*, *Nature* **316**, 817 (1985); D. R. Bell, J. H. Gerlach, N. Kartner, R. N. Buick, V. Ling, *J. Clin. Oncol.* **3**, 311 (1985).
83. E. S. Kawasaki *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5698 (1988).
84. R. Fitts, M. Diamond, C. Hamilton, M. Neri, *Appl. Environ. Microbiol.* **46**, 1146 (1983).
85. J. W. Ward *et al.*, *N. Eng. J. Med.* **318**, 473 (1988).
86. C.-Y. Ou *et al. Science* **239**, 295 (1988); G. J. Murakawa *et al.*, *DNA* 7, 287 (1988).
87. A. Jeffries, V. Wilson, S. Thein, *Nature* **314**, 67 (1985).
88. R. Higuchi, C. H. von Beroldingen, G. F. Sensabaugh, H. A. Erlich, *ibid.* **332**, 543 (1988).
89. A. E. Covone, D. Mutton, P. M. Johnson, M. Adinolfi, *Lancet* **i**, 841 (1984); J. N. Bulmer and P. M. Johnson, *Placenta* 6, 127 (1985); R. Kozma, M. Chapman, Y. W. Loke, P. M. Johnson, M. Adinolfi, *J. Reprod. Immunol.* **11**, 55 (1987).
90. H. Acha-Orbea *et al.*, *Cell* **54**, 263 (1988); J. Urban *et al.*, *ibid.*, p. 577.
91. D. St. Louis and I. M. Verma, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3150 (1988).
92. M. Lemaitre, B. Bayard, B. Lebleu, *ibid.* **84**, 648 (1987).
93. N. A. Holtzman, *Am. J. Hum. Genet.* **42**, 624 (1988).
94. S. L. Comerford, *Biochemistry* **10**, 1993 (1971).
95. Human Gene Mapping Workshop 9, *Cytogenet. Cell Genet.* **46**, 7 (1982); V. A. McKusick, *Mendelian Inheritance in Man* (Johns Hopkins Univ. Press, Baltimore, ed. 7, 1986).
96. The authors gratefully acknowledge support from NSF to U.L., R.K., and L.H. and from NIH, the Muscular Dystrophy Foundation, and Howard Hughes Medical Institute to C.T.C. We thank N. Arnheim, W. Grody, C. Epstein, J. Korenberg, and M. Muldoon for useful advice. C. Höög, T. Hunkapiller, W. Laug, D. Nickerson, and B. Popko offered helpful criticism on the manuscript.